# Comparison of Inverse-Wishart and Separation-Strategy Priors for Bayesian Estimation of Covariance Parameter Matrix in Growth Curve Analysis

Haiyan Liu and Zhiyong Zhang
University of Notre Dame

Kevin J. Grimm
Arizona State University

2017

Author Note

This study was partially supported by a grant from the Department of Education (R305D140037). However, the contents of the study do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

Citation: Liu, H., Zhang, Z., & Grimm, K. J. (2016). Comparison of Inverse-Wishart and Separation-Strategy Priors for Bayesian Estimation of Covariance Parameter Matrix in Growth Curve Analysis. Structural Equation Modeling, 23 (3), 354-367.

## Abstract

Growth curve modeling provides a general framework for analyzing longitudinal data from social, behavioral, and educational sciences. Bayesian methods have been used to estimate growth curve models, in which priors need to be specified for unknown parameters. For the covariance parameter matrix, the inverse-Wishart prior is most commonly used due to its proper and conjugate properties. However, many researchers have pointed out that the inverse-Wishart prior might not work as expected. The purpose of this study is to investigate the influence of the inverse-Wishart prior and compare it with a class of separation-strategy priors on the parameter estimates of growth curve models. This paper first illustrates the use of different types of priors through two real data analyses, and then conducts simulation studies to evaluate and compare these priors in estimating both linear and nonlinear growth curve models. For the linear model, the simulation study shows that both the inverse-Wishart and the separation-strategy priors work well for the fixed effects parameters. For the Level 1 residual variance estimate, the separation-strategy prior performs better than the inverse-Wishart prior. For the covariance matrix, the results are mixing. Overall, the inverse-Wishart prior is suggested if the population correlation coefficient and at least one of the two marginal variances are large. Otherwise, the separation-strategy prior is preferred. For the nonlinear growth curve model, the separation-strategy priors work always better than the inverse-Wishart prior.


*Keywords:* Growth curve models, Bayesian estimation, covariance matrix, inverse-Wishart prior, separation-strategy prior

# Comparison of Inverse-Wishart and Separation-Strategy Priors for Bayesian Estimation of Covariance Parameter Matrix in Growth Curve Analysis

## Introduction

Longitudinal studies are common in social, behavioral and educational sciences. In a longitudinal study, data are collected repeatedly by tracking the same participants over time (e.g., Bock, 1975; Hedeker & Gibbons, 2006; Hsiao, 2003). Through longitudinal data analysis, one can investigate both the intraindividual changes over time and the interindividual differences in the intraindividual changes simultaneously (e.g., Baltes & Nesselroade, 1979).

Many statistical models are available for analyzing longitudinal data, such as repeated-measures ANOVA and growth curve models (e.g., Bollen & Curran, 2006; Hedeker & Gibbons, 2006; Livingston & State, 2012; McArdle, 2009; Singer & Willett, 2003). In recent decades, researchers have found that growth curve models have the advantage of modeling both means and variances and covariances of the initial level and the rate of change simultaneously (e.g., Bryk & Raudenbush, 1987; Raykov, 1993; Rogosa et al., 1982). As a consequence, they have gained popularity in applied research (e.g, McArdle, 1998, 2009; Meredith & Tisak, 1990). In a growth curve model, the "time" variable is usually treated as a continuous predictor and the outcome variable is a function of both time and measurement error. When the means are assumed to be a linear function of time, we have the commonly used linear growth curve model (LGCM, e.g., Lairde & Ware, 1982). Otherwise, a general nonlinear growth curve model may be applied, for instance the logistic growth curve models, Gompertz growth curve models, and Richards growth curve models(e.g., Cameron et al., 2014). In the literature, there are also other variates of growth curve models, for instance, Li et al. (2000) and X. Y. Song et al. (2009) investigated the interaction effects in growth curve models.

Due to their advantages in estimating complex models and the emerging of new software such as BUGS (e.g., Lunn et al., 2012), full Bayesian estimation methods are

increasingly used in growth curve modeling (e.g., Elliott et al., 2005; X. Y. Song & Lee, 2001, 2002; P. Song et al., 2007; Zhang et al., 2007, 2013). Bayesian methods, however, require the explicit specification of prior distributions for parameters to be estimated (e.g., Gelman et al., 2003). Because inverse-Gamma and inverse-Wishart distributions are often proper and conjugate to the Gaussian likelihood, they are the most commonly used priors for a variance parameter or a covariance parameter matrix when data are assumed to follow a univariate or multivariate normal distribution. However, Gelman (2006) was against the use of the inverse-Gamma as a prior distribution for the univariate variance (see also, Gelman et al., 2003). The reason is that the inverse-Gamma distribution has a narrow peak around 0 and thus can be unintentionally informative, which conflicts with the initial purpose of obtaining objective inferences by using such a prior. Other types of priors such as half-t, half-Cauchy, and uniform distributions for the standard deviations were proposed and studied as potentially less informative priors (e.g., Gelman, 2006).

Given that the inverse-Wishart distribution is a multivariate generalization of the inverse-Gamma distribution, it is expected that the inverse-Wishart prior might have the same problems as, or even severer than, the inverse-Gamma prior. Because of its multivariate nature, it is even harder to understand the influence of the inverse-Wishart prior intuitively. If a matrix $\mathbf{M}$ is a sample from the inverse-Wishart distribution $\mathrm{IW}(m, \mathbf{V})$ with the degrees of freedom $m$ and the scale matrix $\mathbf{V}$, its inverse $\mathbf{M}^{-1}$ is from the Wishart distribution $\mathrm{W}(m, \mathbf{V}^{-1})$ and there must be a sequence of random column vectors $\mathbf{x}_1, \mathbf{x}_{2,} \cdots, \mathbf{x}_m \sim \mathrm{MVN}(\mathbf{0}, \mathbf{V})$, where MVN is the short form of "multivariate normal", such that

$$\mathbf{M}^{-1} = \sum_{i=1}^{m} \mathbf{x}_i \mathbf{x}_i^T.$$

As a consequence, $\mathbf{M}^{-1}$ must be non-negative definite and all the diagonal elements have the same degrees of freedom (e.g., Barnard et al., 2000). These restrictions make the components of $\mathbf{M}$ depend on each other. A recent study on the visualization of the inverse-Wishart distribution by Tokuda et al. (2012) found that large correlation

coefficients correspond to large marginal variances in an inverse-Wishart distribution. Therefore, the inverse-Wishart priors might be highly informative, and overwhelmingly influential in the posterior distributions of the covariance matrices. For example, they may cause large bias in parameter estimates, especially when the correlation coefficients are large but marginal variances are small, and vice versa.

Forming new types of priors for covariance matrices can be very difficult. A popular way to form new priors for a covariance matrix is through the matrix decomposition. Barnard et al. (2000) introduced a separation strategy to decompose a covariance matrix $\mathbf{\Psi}$ into a diagonal matrix $\mathbf{S}$ of standard deviations and a correlation matrix $\mathbf{R}$ such that

$$\mathbf{\Psi} = \mathbf{SRS},$$

where $\mathbf{S} = (s_{ij})$ with $s_{ij} \neq 0$ only if $i = j$ and the diagonal element $s_{ii}$ is the standard deviation of the $i$th variable. After decomposition, priors for the elements of $\mathbf{S}$ and $\mathbf{R}$ can be independently specified (e.g., Lunn et al., 2012). Barnard et al. (2000) used the log-normal prior for the vector of standard deviations. For the correlation matrix $\mathbf{R}$ they discussed two types of priors. One is to use a uniform prior for each correlation. The other is the jointly uniform prior $p(\mathbf{R}) \propto 1$. Such priors for the covariance parameter matrix eliminate the dependence among the variance components and correlation coefficients of a covariance matrix, which yet exists in an inverse-Wishart distribution. In addition, due to the structural flexibility of the separation-strategy priors, one can potentially utilize a large variety of priors for the marginal variances such as those used for the univariate variance by Gelman (2006).

In the existing literature on the Bayesian estimation of growth curve models, the majority, if not all, of the studies have directly adopted the inverse-Wishart priors (e.g., Congdon, 2003; Lu et al., 2011; J. H. Pan et al., 2008; Zhang et al., 2013; Zhang & Nesselroade, 2007). However, it is not clear how such priors influence growth curve model

parameter estimates. Furthermore, given Gelman (2006) has shown that the alternative priors for the univariate variance can work better than the default inverse-Gamma distribution, it is important to investigate whether there exists a set of better priors for the covariance matrix based on the separation strategy.

The purpose of this study is to evaluate and compare the performance of the inverse-Wishart prior and the separation-strategy priors on parameter estimates in the framework of latent growth curve modeling. In the following sections, we start with a brief introduction to growth curve models. We then discuss the Bayesian estimation methods and present details on the specification of different types of priors. After that, we first compare the performance of the inverse-Wishart prior and the separation-strategy priors through two real data examples, and then conduct simulation studies to evaluate and compare the performance of the two types of priors in both linear and nonlinear growth curve models. In the end, we discuss the implications and suggestions on the specification of priors in growth curve modeling.

## Growth Curve Models

Growth curve models have been presented in different forms, for instance as structural equation models (SEM, e.g., McArdle & Epstein, 1987), as multilevel models (e.g., Singer & Willett, 2003), and as mixed-effects models (e.g., J. Pan & Fang, 2002).
A growth curve model can be written in the following general form, (I suggested using $\gamma$ instead of $c$ to be consistent with $\beta$)

$$y_{it} = f(t, \boldsymbol{\eta}_i, \boldsymbol{\gamma}) + e_{it}, \tag{1}$$

$$\boldsymbol{\eta}_i = \boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \tag{2}$$

where $y_{it}$ is the observation of person $i$ at time $t$; $e_{it}$ is the intraindividual measurement errors, and the latent variable $\boldsymbol{\eta}_i$ is a vector of growth parameters, which are also called *random effects,* and they vary from person to person to represent the interindividual

differences. The means of the random effects $\boldsymbol{\eta}_i$ are denoted by $\boldsymbol{\beta}$, which are called *fixed effects,* and are the same for all individuals. $\boldsymbol{\epsilon}_i$ is a vector of the residuals of the random effects. $\boldsymbol{\gamma}$ represents the collection of parameters other than $\boldsymbol{\beta}$ that are fixed across individuals. This type of parameters, if they exist, can describe the overall characteristics of the growth trajectories. For instance, they might be the overall lower or upper asymptote of all trajectories. The function $f(t, \boldsymbol{\eta}_i, \boldsymbol{\gamma})$ describes the pattern of each individual's trajectory.

We follow the literature of the growth curve models by assuming that intraindividual measurement errors are identically and independently normally distributed across both individuals and all occasions(e.g., Fitzmaurice et al., 2011),

$$e_{it} \overset{iid}{\sim} \mathrm{N}(0, \sigma_e^2), \tag{3}$$

where $\sigma_e^2$ is an unknown scale parameter, which is also called *Level 1 residual variance.* In addition, the residuals of the growth parameters are also assumed to be identically and independently normally distributed,

$$\boldsymbol{\epsilon}_i \overset{iid}{\sim} \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Psi}) \tag{4}$$

where $\boldsymbol{\Psi}$ is a $q \times q$ covariance matrix when $\boldsymbol{\eta}_i$ is a $q \times 1$ vector.

## Linear Growth Curve Model

Although it is of simple form, the linear growth curve model (LGCM) has been widely used due to its clear interpretation of model parameters. For the linear growth curve model, we have

$$\boldsymbol{\eta}_i = \begin{bmatrix} L_i \\ S_i \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_L \\ \beta_S \end{bmatrix}, \boldsymbol{\Psi} = \begin{bmatrix} \sigma_L^2 & \rho\sigma_L\sigma_S \\ \rho\sigma_L\sigma_S & \sigma_S^2 \end{bmatrix}, \tag{5}$$

where $L_i$ and $S_i$ are the random intercept and random slope associated to individual $i$; and their means are represented by $\beta_L$ and $\beta_S$, which are the same across different individuals; $\boldsymbol{\Psi}$ is the covariance matrix of the random effects and $\sigma_L^2$ and $\sigma_S^2$ are the variance parameters, representing the variability of random intercept and random slope. The correlation coefficient $\rho$ describes the linear relationship between the initial level and the slope.

In the literature, the linear growth trend function $f(\cdot)$ may have different forms (e.g., Preacher et al., 2008). In this study, we take

$$f(t, \boldsymbol{\eta}_i, \boldsymbol{c}) = f(t, \boldsymbol{\eta}_i) = L_i + (t-1)S_i. \tag{6}$$

With this specific form, the random intercept $L_i$ represents the initial level of participant $i$ and $S_i$ represents the rate of change with respect to unit change of time.

## Gompertz Growth Curve Model

Nonlinear growth curve models, such as the Gompertz model, have also been used in the literature. Although the Gompertz growth curve is for long used by researchers to describe the growth processes in both biology and economics(e.g., Winsor, 1932), it is only recently used by psychometricians to represent the growth in human development (e.g., Grimm & Ram, 2009). In our current study, we adopted the specific Gompertz curve model used by Cameron et al. (2014) in which,

$$\boldsymbol{\eta}_i = \begin{bmatrix} b_{i1} \\ b_{i2} \\ b_{i3} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \boldsymbol{\Psi} = \begin{bmatrix} \sigma_1^2 & \rho_1\sigma_1\sigma_2 & \rho_2\sigma_1\sigma_2 \\ \rho_1\sigma_1\sigma_2 & \sigma_2^2 & \rho_3\sigma_2\sigma_3 \\ \rho_2\sigma_1\sigma_2 & \rho_3\sigma_2\sigma_3 & \sigma_3^2 \end{bmatrix}. \tag{7}$$

and the trajectory function has the following specific form

$$f(t, \boldsymbol{\eta}_i, \boldsymbol{\gamma}) = \gamma + b_{i1}\exp[-\exp(b_{i2}(t - b_{i3}))]. \tag{8}$$

Given $\gamma, b_{i1}, b_{i2,}$ and $b_{i3}$, $f(t, \boldsymbol{\eta}_i, \gamma)$ corresponds to a S-shaped curve with $\gamma$ as the lower asymptote for each individual and $\gamma + b_{i1}$ as the upper asymptote for individual $i$. Thus $b_{i1}$ is the possible total change for individual $i$. $b_{i2}$ represents the rate approaching the upper asymptote and $b_{i3}$ is the inflection point at which the shape of the curve changes for individual $i$. In our current study, $\gamma$ is fixed across individuals following Cameron et al. (2014).

## Bayesian Estimation and Prior Specification

Statistical inference in Bayesian analysis is based on the posterior distribution of model parameters. In obtaining the posterior distribution, priors are needed. For the linear growth curve model, the model parameters include the fixed effects parameters $\boldsymbol{\beta}$, the covariance matrix $\boldsymbol{\Psi}$, and the Level 1 residual variance $\sigma_e^2$ and for the Gompertz growth curve model, we also need to consider the lower asymptote parameter $\gamma$ . The presence of the random effects $\boldsymbol{\eta}_i$ makes it difficult to get a relative simple form for the posterior distributions $p(\gamma, \boldsymbol{\beta}, \boldsymbol{\Psi}, \sigma_e^2 | \mathbf{y}_i, i = 1, \cdots, N)$ directly. To overcome the difficulty, the data augmentation algorithm proposed by Tanner & Wong (1987) can be used. We augment the data $\mathbf{y}_i = (y_{it})$ with the random effect $\boldsymbol{\eta}_i$. Using the Bayes' theorem, we obtain

$$p(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Psi}, \sigma_e^2 | \mathbf{y}_i, \boldsymbol{\eta}_i, i = 1, \cdots, N) = \frac{[\Pi_{i=1}^N p(\mathbf{y}_i | \sigma_e^2, \boldsymbol{\eta}_i, \gamma) p(\boldsymbol{\eta}_i | \boldsymbol{\beta}, \boldsymbol{\Psi})] p(\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\Psi})}{p(\mathbf{y}_i, \boldsymbol{\eta}_i, i = 1, \cdots, N)}, \qquad (9)$$

where $[\Pi_{i=1}^N p(\mathbf{y}_i | \sigma_e^2, \boldsymbol{\eta}_i, \boldsymbol{\gamma}) p(\boldsymbol{\eta}_i | \boldsymbol{\beta}, \boldsymbol{\Psi})]$ is the likelihood function; $p(\mathbf{y}_i, \boldsymbol{\eta}_i, i = 1, \cdots, N)$ is the marginal distribution of the augmented data; and $p(\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\Psi})$ is the prior distribution of parameters that is decided before the data collection. By averaging over all possible $\boldsymbol{\eta}_i's$, we can obtain the approximated marginal posterior distributions $p(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Psi}, \sigma_e^2 | \mathbf{y}_i, i = 1, \cdots, N)$. However, the distribution of $\boldsymbol{\eta}_i's$ in turn depends on $(\boldsymbol{\beta}, \boldsymbol{\Psi})$. We thus can use the Markov Chain Monte Carlo (MCMC) algorithms to get samples of both $(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Psi}, \sigma_e^2)$ and $\boldsymbol{\eta}_i$ from their conditional posterior distributions(e.g., Robert & Casella, 2004).

As seen from the posterior distribution in Equation (9), the prior distribution $p(\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\Psi})$ is required and it influences the posterior inference of parameters. As a result, it is important to choose priors in Bayesian analysis. For convenience, it is usually assumed that the prior knowledge on the parameter $\boldsymbol{\gamma}$, the fixed effects $\boldsymbol{\beta}$, the Level 1 residual variance $\sigma_e^2$, and the covariance matrix $\boldsymbol{\Psi}$ are independent, so that

$$p(\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\Psi}) = p(\boldsymbol{\gamma})p(\boldsymbol{\beta})p(\sigma_e^2)p(\boldsymbol{\Psi}).$$

To reduce the influence of priors, researchers often prefer non-informative priors even though Bayesian methods allow the incorporation of prior information (e.g., Zhang et al., 2007). Therefore, in this study, we focus on the use of non-informative priors.

## Priors for $\boldsymbol{\gamma}, \boldsymbol{\beta}$

Both $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are fixed for all individuals. Their priors are usually easier to specify than $\sigma_e^2$ and $\boldsymbol{\Psi}$. For the rest of discussion, we adopt independent normal prior $\mathrm{N}(0, 10^{-4})$ for each element in $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. The priors for $\sigma_e^2$ and $\boldsymbol{\Psi}$ are specified soon after wards.

## Priors for $\sigma_e^2$

The inverse-Gamma (IG) prior is most widely used for $\sigma_e^2$ although other priors have been recommended. An inverse-Gamma distribution, $\mathrm{IG}(\alpha, \delta)$ has the density function

$$p(x; \alpha, \delta) = \frac{\delta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\delta}{x}\right), \tag{10}$$

where $\alpha$ is the shape parameter and $\delta$ is the the scale parameter. To reduce the information in an inverse-Gamma prior, small $\alpha$ and $\delta$ are preferred. Recently, Gelman (2006) has recommended the use of the half-t distribution for the standard deviation parameter $\sigma_e$. As a special case of half-t family, the half-Cauchy (HC) distribution has been intensively studied by Polson & Scott (2012). A half-Cauchy distribution with mean 0

and scale $\tau$ has the density function

$$p(x,\tau) = \frac{2}{\pi} \cdot \frac{\tau}{x^2+\tau^2}, \tag{11}$$

and its amplitude is $\frac{2}{\pi\tau}$. Geometrically, $\tau$ is the scale parameter which specifies the half-width at half-maximum, i.e, $p(\tau,\tau) = \frac{1}{\pi\tau}$. Therefore, a larger $\tau$ leads to a lower but wider peak around the origin, and thus less informative. The Cauchy distribution is a distribution of the ratio of two independently normally distributed random variables. Therefore, one can sample from Cauchy$(0,\tau)$ by obtaining the ratio of samples of two independent normal distributions $\mathrm{N}(0,\tau^2)$ and $\mathrm{N}(0,1)$. Gelman (2006) used $\tau = 25$. Another special distribution from the half-t family is the non-negative uniform distribution

$$p(x) = \mathrm{U}[0,\infty). \tag{12}$$

Compared to the inverse-Gamma distribution, the half-Cauchy distribution has less mass near the origin and can have a heavier tail. Compared to the uniform distribution, the half-Cauchy distribution favors finite variances, which is more meaningful in practice. Therefore, in this study, we use the half-Cauchy distribution HC(0, 25) as the prior for $\sigma_e$ under all conditions to focus on the evaluation of the priors for the covariance matrix.

## Priors for $\boldsymbol{\Psi}$

Two types of priors are used for the covariance parameter matrix $\boldsymbol{\Psi}$: the inverse-Wishart prior and the separation-strategy prior. For the separation-strategy prior, we further consider three different specifications as discussed below.

**The inverse-Wishart prior.** The inverse-Wishart distribution $\mathrm{IW}(m,\mathbf{V})$ with the degrees of freedom $m$ and the scale matrix $\mathbf{V}$ is the most widely used prior for the covariance matrix $\boldsymbol{\Psi}$. This is mainly because for the Gaussian likelihood, $\mathrm{IW}(m,\mathbf{V})$ is a conjugate prior for the covariance matrix (e.g., Gelman et al., 2003). Therefore, the

posterior distribution for the covariance matrix still belongs to the inverse-Wishart family. The density function of $\mathrm{IW}(m, \mathbf{V})$ is

$$f(\mathbf{\Psi}|m, \mathbf{V}) = \frac{|\mathbf{V}|^{\frac{m}{2}}}{2^{\frac{mq}{2}}\Gamma_q(\frac{m}{2})} |\mathbf{\Psi}|^{-\frac{m+q+1}{2}} e^{-\frac{1}{2}\mathrm{tr}(\mathbf{V}\mathbf{\Psi}^{-1})}, \tag{13}$$

where $q$ is the dimension of covariance matrix $\mathbf{\Psi}$ and $\Gamma_q$ is the multivariate gamma function. In the linear growth curve model, $q = 2$ and in the Gompertz growth curve model, $q = 3$. To use least information in the inverse-Wishart prior, one usually sets $m = q$ (e.g., Congdon, 2003).

**The separation-strategy priors.** For the separation-strategy priors, we specify independent priors to each marginal variance of random effects and their correlation coefficients, which is also suggested by Lunn et al. (2012). In this study, we use a uniform prior for the correlation coefficients $\rho$,

$$p(\rho) = \mathrm{U}[-1, 1] = \frac{1}{2}.$$

where $\rho$ could be any correlation coefficients in the covariance matrix $\mathbf{\Psi}$.

Because previous studies have suggested that different priors for the variance parameter can be used (e.g., Gelman, 2006; Polson & Scott, 2012), in our current study, we investigate three priors for marginal variances as discussed below.

*SS1 prior.* For all the marginal variances, the identical and independent inverse-Gamma priors $\mathrm{IG}(10^{-4}, 10^{-4})$ are used.

*SS2 prior.* Instead of specifying priors directly for $\sigma_L^2$ and $\sigma_S^2$, $\sigma_1^2, \sigma_2^2, \sigma_3^2$, we use the independent uniform prior for the standard deviations, $p(x) = \mathrm{U}[0, \infty)$, where $x =\sigma_L$, $\sigma_S$, $\sigma_1$, $\sigma_2$, or $\sigma_3$.

*SS3 prior.* In this specification, the half-Cauchy $\mathrm{HC}(0, 25)$ prior is used for both $\sigma_L$ and $\sigma_S$, $\sigma_1, \sigma_2$, and $\sigma_3$.

## Real Data Analysis Examples

To illustrate the use of the inverse-Wishart and the separation-strategy priors, we apply them in the analysis of the subsets of data on Wechsler Intelligence Scale for Children (WISC) [1] and the Early Childhood Longitudinal Study-Kindergarten Cohort (ECLS-K).

### Linear modeling of WISC data

The data used here include scores on 204 school children who were measured 4 times on his/her verbal ability at grades 1, 2, 4, and 6, which corresponds to $t = 0, 1, 3, 5$. Both the trajectory plot and previous data analysis (e.g., McArdle & Nesselroade, 2014) suggested that a linear growth curve model is plausible for the current data and, therefore, we fit the linear growth curve model to the data. Four sets of priors, as listed in Table 1, are used in the analysis. Note that the same priors are used for $\sigma_e$, $\beta_L$, and $\beta_S$. For the covariance matrix, both the inverse-Wishart prior and the three separation-strategy priors are used. The separation-strategy priors are different in terms of the prior choice for $\sigma_L$ and $\sigma_S$.

Table 2 compares the Bayesian parameter estimates based on the four sets of priors as well as the maximum likelihood estimates (MLE). To get the Bayesian estimates, a total of 120,000 iterations are used with the first 80,000 iterations discarded as the burn-in period. The kept Markov chain for each parameter passed the Geweke test of convergence and eye-ball checking of the history plot (e.g., Gelman et al., 2003; Geweke, 1992). To evaluate the influence of the priors, the parameter estimates from the Bayesian method are compared with those from MLE. In particular, we define a bias measure as the percentage of the difference between the Bayesian estimates and MLE over MLE.

From Table 2, the use of all four types of priors gives similar estimates of the fixed effects with bias less than 1% and similar standard deviations. For the Level 1 residual variance, variances of the slopes, and the correlation between slope and intercept, the inverse-Wishart prior appears to lead to larger bias than the separation-strategy priors.

[1] We thank X X for allowing us to use the data.

Particularly, the use of the inverse-Wishart prior underestimates the variances of the random effects but overestimates the correlation coefficient. The inverse-Wishart prior causes large bias ($> 10\%$) on the correlation coefficient, however the separation-strategy priors do not. In this practical example, the correlation coefficient describes the linear relationship between the initial level and the rate of change of the verbal ability. The squared correlation coefficient thus represents the proportion of the variability existing in the random rate of change that can be attributed to the variability of the initial level of children's verbal ability. Hence an accurate estimate of the correlation coefficient would be of particular interest to researchers.

## Nonlinear modeling of ECLS-K data

Data used here are from 500 children whose math achievement was measured between age 5 and 14. Math scores were collected for each kids in the Fall and Spring semesters of Kindergarten and 1st grade, , as well as the Spring semesters of 3rd, 5th, and 8th grades, which are coded as $t = 0, 0.5, 1, 1.5, 3.5, 5.5, 8.5$. We fitted the Gompertz curve model (7)and (8) to the ECLS-K data as suggested by Cameron et al. (2014), but estimated the parameters in the Bayesian framework. The prior distributions are similar to what we have used for the linear growth curve model in Table 1. Additionally, $\mathrm{N}(0, 10^4)$ is used as the prior for the lower asymptote parameter $\gamma$. During the analysis, the Gibbs sampling procedure encounter some problems. Some of the sampled covariance matrices are not invertible. This might due to the extremely large sample of correlation coefficients, thus we constrained the priors used for the correlation coefficients and let $\rho_1, \rho_2, \rho_3 \overset{iid}{\sim} \mathrm{U}[-0.95, 0.95]$. In addition, when using the SS2 prior, the sign of the estimates of $\beta_1$ is negative. Recall that $\beta_1$ is mean of total change of math ability and the trajectory plot indicates that it should be positive. Thus, we use the truncated prior $\mathrm{N}(0, 10^4)\mathrm{I}(0, \infty)$ instead of the weak informative prior $\mathrm{N}(0, 10^4)$.

The parameter estimates, standard deviations, and Geweke test statistics are

summarized in Table 3. Because we do not have the exact MLE, we are not able to compare the performance of Bayesian estimation methods against the MLE methods. Same as in the linear growth model, a total of 120,000 iterations are used for the Gompertz model and the first 80,000 iterations are discarded as burn-in. With the remaining 40,000 iterations, all the chains passed the Geweke test of convergence. Clearly, the use of the separation-strategy priors results in both similar parameter estimates and standard deviations. However, the estimates with the inverse-Wishart prior are quite different from those with separation-strategy priors. Because, we do not know the underlying parameter value, we cannot conclude which type of priors gives more reliable estimates yet. Therefore, we are going to compare different types of priors through simulation studies.

## Simulation Study I: A linear growth model

In the previous section, we have demonstrated the potential influence of the inverse-Wishart and the separation-strategy priors in the growth curve analyses empirically through the analysis of two sets of real data. To better compare the inverse-Wishart prior with the separation-strategy priors, we conduct two simulation studies on the linear and Gompertz growth curve model, respectively. The first simulation study presented here use the linear growth curve model in the analysis of the WISC data as the population model. The simulation conditions, evaluation criteria, and simulation results for the linear model are presented below.

### Simulation Conditions

A major goal of a longitudinal study is to detect the interindividual differences in intraindividual change, reflected by the variance of the slope (e.g., Singer & Willett, 2003). Therefore, we fix $\beta_L = 20$, $\beta_S = 5$, and $\sigma_L^2$ =20, similar to the estimates in real data analysis (Table 2). We then vary the following factors including the variance of the slope, the correlation between the intercept and slope, the Level 1 residual variance, and the

sample size.[2]

**The Variance of the Random Slope.** The magnitudes of the variance of the random slope influence the power of longitudinal data analysis. The power to detect individual differences in slope is greater when the slope variance is larger (e.g., Hertzog et al., 2008). In addition, Hertzog et al. (2008) concluded that the ratio of the slope and intercept variances was small to moderate in empirical studies (e.g., Hertzog & Schiae, 1986; Lovden et al., 2004). More recently, Ke & Wang (2014) suggested that the ratio was usually less than 1 : 4 in practice. In the simulation, the random intercept variance is fixed at 20 and $\sigma_S^2$ is set at 1, 3, and 5, respectively.

**The Correlation between Intercept and Slope.** In the real data analysis (Table 2), we found notable difference in the estimates of the correlation coefficient of the intercept and slope when using the two types of priors. Furthermore, Takuda et al. (2012) showed that large correlation coefficients are accompanied by large marginal variances statistically. Therefore, one would expect the correlation between the random intercept and slope to play a role in the analysis. In the real data analysis, the correlation estimate is around 0.56, and therefore we consider three levels of correlation $\rho = 0$, 0.5, and 0.8, indicating no correlation, correlation close to the real data analysis, and large correlation.

**Level 1 Residual Variance.** The Level 1 residual variance has been found to influence both power and Type I error of a longitudinal study (e.g, Hertzog et al., 2006, 2008; Ke & Wang, 2014). In the simulation, we set $\sigma_e^2 = 20$ and 5, either greater or smaller than that from the real data analysis (Table 2).

**Number of Participants.** In Bayesian analysis, the posterior inference is a balance between data and priors. Therefore, the influence of the priors is greater when the sample size is smaller. In the real data analysis, the sample size is 204. In the simulation, we consider three levels of sample sizes at $N = 50$, 100, and 200.

[2]Although we use four measurement occasions in the study, the number of occasions does not influence our conclusions on the comparison of the two types of priors.

**Priors.** The four sets of prior used in the real data analysis (Table 1) are also used in the simulation study.

Based on the factorial design, we consider $3 \times 3 \times 2 \times 3 \times 4 = 216$ different conditions in our simulation. Under each condition, 500 replications of data with 4 measurement occasions are generated and analyzed.

## Evaluation Criteria

Let $\theta$ be an arbitrary parameter in the model to be estimated and also its population value. Let $\hat{\theta}_r$ be the estimate of $\theta$ and $[L_r, R_r]$ be the 95% percentile credible interval from the $r$th ($r = 1, 2, \ldots, 500$) simulation replication. In assessing the the performance of the priors, two criteria are used. The first criterion is the bias or relative bias (BIAS), which is defined as

$$\text{BIAS} = \begin{cases} 100 \times \bar{\hat{\theta}} & \theta = 0 \\ 100 \times \frac{\bar{\hat{\theta}} - \theta}{\theta} & \theta \neq 0 \end{cases}, \tag{14}$$

where

$$\bar{\hat{\theta}} = \frac{1}{500} \sum_{r=1}^{500} \hat{\theta}_r. \tag{15}$$

The BIAS quantifies the accuracy of the parameter estimates. Based on Muthén & Muthén (2002), BIAS less than 5% is *ignorable*, BIAS between 5% and 10% means *moderately biased*, and BIAS above 10% is *significantly biased*.

The second criterion is the 95% credible interval coverage rate (CR):

$$\text{CR} = 1 - \frac{\sum_{i=1}^{500} [I_{\{L_r > \theta\}} + I_{\{R_r < \theta\}}]}{500}, \tag{16}$$

where $I_{\{\cdot\}}$ is the indicator function. If there are $R$ independent replications, according to

the Central Limit Theorem,

$$\text{CR} \overset{\mathcal{L}}{\rightsquigarrow} \text{N}(0.95, \frac{0.95 \times 0.05}{R}).$$

Hence, a CR that falls in the range $[0.95 - 1.96\sqrt{0.95 \times 0.05/R}, 0.95 + 1.96\sqrt{0.95 \times 0.05/R}]$ can be considered as an indication of good coverage. In our simulation, $R = 500$, the range should be about $[0.93, 0.97]$. For the convenience of comparison, instead of CR, we report the discrepancy of the coverage rate from 0.95. The discrepancy is defined as

$$\text{DCR} = \text{CR} - 0.95.$$

A CR falling out of the interval $[0.93, 0.97]$ is equivalent to a DCR$> 0.02$ or DCR$< -0.02$. Besides, a greater absolute value of DCR indicates a worse coverage rate.

## Results

Representative results from our simulation are provided in Table 4 through Table 7.[3] In the following, we evaluate the influence of priors on the fixed effects parameters, the Level 1 residual variance, and the covariance matrix of the random effects, respectively, in terms of the relative bias and discrepancy of coverage rate.

**Fixed-effects Parameters $\beta_L, \beta_S$.** Table 4 includes the results for the fixed effects $\beta_L$ and $\beta_S$ when the Level 1 residual variance $\sigma_e^2 = 20$ and the sample size $N = 50$. The relative bias of the fixed effects for all 4 sets of priors falls within the interval $[-1\%, 1\%]$ and the bias is, therefore, ignorable. The majority of DCRs are in the range of $[-0.02, 0.02]$, with three exceptions that are 0.03 (bold numbers in the table). For the scenarios with $\sigma_e^2 = 5$ and $N = 100$, 200, even better performance was observed. Overall, all four sets of priors appear to perform equally well and have limited influence on the estimates of the fixed effects parameters.

[3]Due to limited space, we cannot include all results. Interested readers may find out the complete simulation results on our website.

**Level 1 Residual Variance $\sigma_e^2$.** The BIAS and DCR for $\sigma_e^2$ when its population value is 20 are provided in Table 5. Notably, the sample size plays an important role and when the sample size increases, the bias decreases. This is well expected since the effect of prior decreases with the increases of sample size. Therefore, we compare the four priors for a given sample size. Overall, the separation-strategy priors have less bias than the inverse-Wishart prior. Among the three types of separation-strategy priors, the biases with SS2 and SS3 are close to each other and smaller than that of SS1. The separation-strategy priors have slightly better coverage rate than the inverse-Wishart prior, and overall the inverse-Wishart prior underestimates the coverage rate.

The bias varies with respect to the population values of $\rho$ and $\sigma_S^2$. The bias decreases as the population correlation $\rho$ of the two random effects increases or the population variance of the random slope $\sigma_S^2$ increases. This pattern is especially clear with the separation-strategy priors. Because we used the same priors for $\sigma_e^2$ and the fixed effects, the differences in the estimates of $\sigma_e^2$ should be caused by the priors of the covariance matrix. Therefore, the inverse-Wishart prior exerts a bigger influence on the estimates of $\sigma_e^2$ than the separation-strategy priors, especially SS2 and SS3.

**Covariance Matrix $\boldsymbol{\Psi}$ $(\sigma_L^2, \sigma_S^2, \rho)$ .** The results for the covariance matrix $\boldsymbol{\Psi}$ are provided in Tables 6–7. Table 6 contains the relative bias of $\sigma_L^2, \sigma_S^2$, and $\rho$ when the true Level 1 residual variance $\sigma_e^2 = 20$. When the sample size increases, the bias becomes clearly smaller regardless of the priors. When other factors are fixed but the variance of the random slope $\sigma_S^2$ increases from 1 to 3, then to 5, the performance of the separation-strategy priors is improved with smaller bias. However, this is not the case for the inverse-Wishart prior, which actually reflects the informative property of the inverse-Wishart prior.

The difference between the inverse-Wishart prior and the separation-strategy priors varies according to the magnitudes of the population correlation coefficient between the two random effects. When the population correlation coefficient is 0 and 0.50, the

separation-strategy priors have better estimates than the inverse-Wishart prior. Overall, the bias with the separation-strategy priors is smaller than that with the inverse-Wishart prior, and this pattern is even more clearer when the sample size is as large as 100 and 200.

When the population correlation coefficient is 0.80, the comparison is a bit more complicated. With the sample size 50, bias with the inverse-Wishart prior is smaller than that with the separation-strategy priors. With sample size 100, the bias with the inverse-Wishart prior is smaller when $\sigma_S^2 = 1$ and 3. Furthermore, with the sample size 200, only when $\sigma_S^2 = 1$, the inverse-Wishart prior has smaller bias. As expected, when the sample size is larger, the difference between the inverse-Wishart prior and the separation-strategy priors disappears. In addition, when the true correlation coefficient is 0.80 and $\sigma_S^2 = 1$, the inverse-Wishart prior has smaller bias on the marginal variance of the random intercept and correlation of the two random effects, but relatively larger bias on the marginal variance of the random slope.

Overall, the use of the inverse-Wishart prior tends to underestimate marginal variances but overestimate the correlation coefficients when the population correlation coefficient between the two random effects is 0 or 0.50. While when the population correlation is 0.8 and $\sigma_S^2 = 1$, the inverse-Wishart prior overestimates small marginal variances but underestimates the correlation coefficient. The principle that drove this phenomena will be discussed through the visualization plot of the inverse-Wishart prior IW(2, $\mathbf{I}_{2\times2}$) in Figure 1.

Comparing the three separation-strategy priors, we find that SS2 and SS3 lead to similar bias on the parameter estimates of the covariance matrix, namely, larger bias in estimating the marginal variances but smaller bias in estimating the correlation coefficient than SS1. Recall that in SS2 and SS3, the uniform and half-Cauchy prior for the standard deviations of the marginal variances are used and both priors belong to the t-distribution family and were suggested by Gelman (2006) for the univariate variance. However, our results show that they do not necessarily perform better than the inverse-Gamma prior in

higher dimensional situations.

Table 7 shows the discrepancy of coverage rates (DCR) when the Level 1 residual variance $\sigma_e^2 = 20$. Overall, the separation-strategy priors have DCR closer to 0, which indicates better coverage rate. When the population $\rho$ is as large as 0.80, the use of all four priors leads to bad coverage rate for $\rho$.

## Simulation Study II: Gompertz growth model

In the previous simulation study, we focused on a linear model. In this section, we focus on the Gompertz model used in the ECLS-K data analysis. To generate data, we set $c = 0.15$, $b_1 = 2.80$, $b_2 = 0.46$, $b_3 = 1.56$; $\sigma_e^2 = 0.023$, $\sigma_1^2 = 0.126$, $\sigma_2^2 = 0.007$, $\sigma_3^2 = 0.285$, which are close to the parameter estimates from the ECLS-K analysis. In our previous study on the linear growth curve model, we notice that the relation between the correlation coefficients and the marginal variances influenced the relative performance of the two types of priors. Therefore, we evaluate two sets of correlation coefficients: $(\rho_1, \rho_2, \rho_3) = (0, 0, 0)$, which indicates no correlations and $(0.60, -0.50, -0.80)$, which is from real data. Sample sizes are set at $N = 200$ and 500. The priors used in the simulation are the same as in the analysis of ECLS-K data.

Same as simulation study I, 500 data sets are generated and estimated under each condition using all four groups of priors. The relative biases(14) and discrepancy of coverage rates(16) are summarized in Table 8 and Table 9.

From Table 8 and Table 9, the inverse-Wishart prior $\mathrm{IW}(3, \mathbf{I}_{3\times 3})$ does not work well with extremely large bias and poor coverage rate under all four conditions. The three separation-strategy priors on the other hand have both negligible bias and the discrepancies of the coverage rate of all parameters fall mostly in the interval $[-0.02, 0.02]$, indicating good coverage rates.

## Discussion and Conclusion

Latent growth curve modeling is a commonly used technique to analyze longitudinal data. With the increasing complexity of the model, Bayesian methods are more and more widely used to conduct growth curve analysis (e.g., Lu et al., 2011; Zhang, 2013). In Bayesian analysis, a prior can influence the parameter estimates dramatically especially when the sample size is small. In this paper, we investigated the influence of the inverse-Wishart prior and three separation-strategy priors on the estimates of the covariance matrix. We first demonstrated the effects of the priors in estimating both linear and nonlinear growth curve models through real data analyses. We then conducted two Monte Carlo simulation studies to further evaluate and compare the performance of the four different priors.

The inverse-Wishart prior and the separation-strategy prior are two ways to specify priors for the same covariance parameter matrix. In an inverse-Wishart prior, a covariance matrix is treated as an entity. When we use an inverse-Wishart prior, the marginal variances and covariances are taken as parts of the matrices sampled from an inverse-Wishart distribution. The sampled matrices automatically satisfy the restrictions such as non-negative definite and same degrees of freedom of the marginal variances (e.g., Barnard et al., 2000). However, in a separation-strategy prior, there is no such dependence among the prior knowledge of the components of $\mathbf{\Psi}$. Besides, the marginal variances do not need to share the same degree of freedom as that in a matrix from an inverse-Wishart distribution.

In our current study, we investigate on the priors distributions of covariance matrix parameters of sizes 2 by 2 and 3 by 3 and in the contexts of both linear and nonlinear growth curve models, respectively. Through the simulation studies, we find that overall the separation-strategy priors perform better than the inverse-Wishart prior in the estimation of both linear and nonlinear growth curve models. The estimates with the separation-strategy priors have both smaller biases and better coverage rates. Therefore,

we recommend the use of separation-strategy priors in overall.

For linear growth curve models, there might be some exceptions. The inverse-Wishart priors might be preferred if we "believe" both of the true marginal variances and the correlation coefficients of the random effects are large. Figure 1 contains two plots about the inverse-Wishart distribution. The left-panel is the scatter plot of the first marginal variances and the correlation coefficients of covariance matrices from the inverse-Wishart distribution IW(2, $\mathbf{I}_{2\times 2}$) and the right panel is the approximated density plot of the correlation coefficients. From the right panel of the plot, we can notice that the marginal distribution of the correlation coefficient $\rho$ is not uniform but favors values close to $-1$ and 1. From the left panel, we can observe that the large correlation coefficient corresponds to the large marginal variance on average. Hence, in the inverse-Wishart prior, the implied marginal variance and correlation coefficient tends to be large. If the population parameters adopt the pattern indicated by the inverse-Wishart distribution, the overall performance of such a prior will be beneficial. However, in practice, one can hardly know the parameter values without specifying priors first. Therefore, one can conduct a sensitivity analysis to evaluate how model parameter estimates differ according to different priors (e.g., Gelman et al., 2003)

For the Gompertz model, the separation-strategy priors work consistently better than the inverse-Wishart$(3, \boldsymbol{I}_{3\times 3})$. With the separation-strategy priors, the parameter estimates have both negligible biases and good coverage rates. However, with the inverse-Wishart prior$(3, \mathbf{I}_{3\times 3})$, the biases are surprisingly large and the coverage rates are very poor. Although we could incorporate extra information in choosing the prior distribution and use alternative scale matrix for the inverse-Wishart prior, it is very hard in practice. This is probably why in the current literature researchers very often use the identity scale matrix for the inverse-Wishart priors(e.g., Cohen et al., 2003; Ghosh & Dunson, 2009; J. H. Pan et al., 2008; Zhang, 2013).

Although we have focused on both linear and Gompertz growth curve models, the

method can be extended to other models. In practice, with the increase of the dimension of covariance matrices, the use of separation-strategy priors might cause some practical issues. For example, the singularity of covariance matrix might be one of the major problems we may encounter. Furthermore, Bayesian estimation with separation-strategy priors take much longer time than with inverse-Wishart priors to obtain posterior samples. It is thus very costly to perform a simulation study.

In social, behavioral, and education sciences, covariance structures are of great interests to researchers. In the existing literature, almost all studies have applied the inverse-Wishart prior in Bayesian estimation. We hope our study can draw attention to the choice of priors on the covariance matrices in the future.

# References


Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). New York: Academic Press.

Barnard, J., McCulloch, R., & Meng, X. (2000). Modeling covariances matrices in terms of standard deviations and correlations with applications to shrinkage. *Statistica Sinica*, *10*, 1281–1311.

Bock, R. D. (1975). Basic issues in the measurement of change. In D. N. De Gruijter & L. J. van der Kamp (Eds.), *Advances in psychological and educational measurement* (pp. 75–96). New York: John Wiley and Sons.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. New Jersey: Wiley.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*, 147–158.

Cameron, C. E., Grimm, K., Steele, J. S., & Castro-Schilo, L. (2014). Nonlinear gompertz curve models of achievement gaps in mathematics and reading. *Journal of Educational Psychology*.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Hillsdale NJ: Lawrence Erlbaum Associates.

Congdon, P. (2003). *Applied Bayesian modelling*. New York: John Wiley & Sons, Inc.

Elliott, M. R., Gallo, J. J., Ten Have, T. R., Bogner, H. R., & Katz, I. R. (2005). Using a bayesian latent growth curve model to identify trajectories of positive affect and negative events following myocardial infarction. *Biostatistics*, *6*(1), 119-143.

Fitzmaurice, G., Laird, N., & Ware, J. (2011). *Applied longitudinal analysis.* New Jersey: John Wiley and Sons, Inc.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*(3), 515–533.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian Data Analysis.* London, UK: Chapman & Hall/CRC.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernado, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics 4* (pp. 169–193). Oxford, UK: Clarendon Press.

Ghosh, J., & Dunson, D. B. (2009). Bayesian model selection in factor analytic models. In D. Dunson (Ed.), *Random effect and latent variable model selection.* John Wiley & Sons.

Grimm, K., & Ram, N. (2009). Nonlinear growth models in mplus and sas. *Structural Equation Modeling*, *16*, 676-701.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Hertzog, C., Lindenberger, U., Ghisletta, P., & von Oertzen, T. (2006). On the power of multivariate latent growth curve models to detect correlated change. *Psychological Methods*, *11*(3), 244–252.

Hertzog, C., & Schiae, K. W. (1986). Stability and change in adult intelligence: 1. Analysis of longitudinal covariance structures. *Psychology and Aging*, *1*(2), 159–171.

Hertzog, C., von Oertzen, T., Ghisletta, P., & Lindenberger, U. (2008). Evaluating the power of latent growth curve models to detect individual differences in change. *Structural Equation Modeling*, *15*(3), 541–563.

Hsiao, C. (2003). *Introduction to the theory of neural computation* (2nd ed.). London, UK: Cambridge University Press.

Ke, Z., & Wang, L. (2014). *Detecting individual differences in change: Methods and comparisons.* (To appear in Structure Equation Modeling: A multidisciplinary Journal)

Lairde, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963–974.

Li, F. Z., Duncan, T. E., & Acock, A. (2000). Modeling interaction effects in latent growth curve models. *Structure Equation Modeling: A Multidisciplinary Journal*, *7*(4), 497-453.

Livingston, M. A., & State, A. (2012). Selecting a linear mixed model for longitudinal data: Repeated measures analysis of variance, covariance pattern model, and grwoth curve approaches. *Psychology Methods*, *17*(1), 15-30.

Lovden, M., Ghisletta, P., & Lindenberger, U. (2004). Path coefficients and path regressions alternative complemetart concepts. *Aging, Newropsychology, and Cognition*(189–202).

Lu, Z., Zhang, Z., & Lubke, G. (2011). Bayesian inference for growth mixture models with non-ignorable missing data. *Multivariate Behavioral Research*, *46*(4), 567–597.

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis.* Boca Raton, FL: Chapman & Hall/CRC.

McArdle, J. J. (1998). Modeling longitudinal data by latent growth curve methods. In G. Marcoulides (Ed.), *Modern methods for business research* (pp. 359–406). Mahwah, NJ: Lawrence Erlbaum Associates.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *33*(60), 577–605.

McArdle, J. J., & Epstein, D. B. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*(1), 110–133.

McArdle, J. J., & Nesselroade, J. (2014). *Longitudinal data analysis using structural equation models*. Washington, D.C.: American Psychological Association.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107–122.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*(4), 599–620.

Pan, J., & Fang, K. (2002). *Growth curve models and statistical diagnostics*. New York: Springer.

Pan, J. H., Song, X. Y., Lee, S. Y., & Kwork, T. (2008). Longitudinal analysis of quality of life for stroke survivors using latent curve models. *Stroke*, *39*, 2795-2802.

Polson, N., & Scott, J. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis*, *7*(2), 1–16.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*. Los Angeles: Sage.

Raykov, T. (1993). A structural equation model for measuring residualized change and discerning patterns of growth of decline. *Applied Psychological Measurement*, *17*(1), 53–71.

Robert, C. P., & Casella, G. (2004). *Monte Carlo statistical methods*. New York: Springer.

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, *92*, 726-748.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. London, UK: Oxford Universal Academy Press.

Song, P., Zhang, P., & Qu, A. (2007). Maximum likelihood inference in robust linear mixed-effects models using multivariate t distribution. *Statistica Sinica*, *17*, 929–943.

Song, X. Y., & Lee, S. Y. (2001). Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations. *British Journal of Mathematical and Statistical Psychology*, *54*, 237–263.

Song, X. Y., & Lee, S. Y. (2002). Bayesian estimation and model selection of multivariate linear model with polytomous variables. *Multivariate Behavioral Research*, *37*, 453-477.

Song, X. Y., Lee, S. Y., & Hser, Y. I. (2009). Bayesian analysis of multivariate latent curve models with nonlinear longitudinal latent effects. *Structure Equation Modeling: A Multidisciplinary Journal*, *16*(2), 245-266.

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the american statistical association*, *82*, 528–540.

Tokuda, T., Goodrich, B., Mechelen, I., Gelman, A., & Tuerlinckx, F. (2012). *Visualizing distributions of covariance matrices.* (http://www.stat.columbia.edu/∼gelman/research/unpublished/Visualization.pdf)

Winsor, C. (1932). The gompertz curve as a growth equation. *Proc Natl Acad Sci U S A*, *18*(1), 1-8.

Zhang, Z. (2013). Bayesian growth curve models with the generalized error distribution. *Journal of Applied Statistics*, *40*(8), 1779-1795.

Zhang, Z., Hamagami, F., Wang, L., Grimm, K. J., & Nesselroade, J. R. (2007). Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development*, *31*(4), 374–383.

Zhang, Z., Lai, K., Lu, Z., & Tong, X. (2013). Bayesian inference and application of robust growth curve models using Student's t distribution. *Structural Equation Modeling*, *20*, 47–78.

Zhang, Z., & Nesselroade, J. R. (2007). Bayesian estimation of categorical dynamic factor models. *Multivariate Behavioral Research*, *42*(4), 729-756.

Table 1
*Priors used in the analysis of the WISC data*

| IW | SS1 | SS2 | SS3 |
|---|---|---|---|
| | $\sigma_L^2 \sim \mathrm{IG}(10^{-4}, 10^{-4})$ | $\sigma_L \sim \mathrm{U}[0, \infty)$ | $\sigma_L \sim \mathrm{HC}(0, 25)$ |
| $\boldsymbol{\Psi} \sim \mathrm{IW}(2, \mathbf{I}_{2\times 2})$ | $\sigma_S^2 \sim \mathrm{IG}(10^{-4}, 10^{-4})$ | $\sigma_S \sim \mathrm{U}[0, \infty)$ | $\sigma_S \sim \mathrm{HC}(0, 25)$ |
| | $\rho \sim \mathrm{U}[-1, 1]$ | $\rho \sim \mathrm{U}[-1, 1]$ | $\rho \sim \mathrm{U}[-1, 1]$ |
| $\sigma_e \sim \mathrm{HC}(0, 25)$ | $\sigma_e \sim \mathrm{HC}(0, 25)$ | $\sigma_e \sim \mathrm{HC}(0, 25)$ | $\sigma_e \sim \mathrm{HC}(0, 25)$ |
| $\boldsymbol{\beta} \sim \mathrm{MVN}(\mathbf{0}, 10^4\mathbf{I}_{2\times 2})$ | $\boldsymbol{\beta} \sim \mathrm{MVN}(\mathbf{0}, 10^4\mathbf{I}_{2\times 2})$ | $\boldsymbol{\beta} \sim \mathrm{MVN}(\mathbf{0}, 10^4\mathbf{I}_{2\times 2})$ | $\boldsymbol{\beta} \sim \mathrm{MVN}(\mathbf{0}, 10^4\mathbf{I}_{2\times 2})$ |

Table 2

*Parameter estimates of the linear growth curve analysis of the WISC data.*

| | Estimate | BIAS | | | | SD | | | | Geweke statistic | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Par | ML | IW | SS1 | SS2 | SS3 | IW | SS1 | SS2 | SS3 | IW | SS1 | SS2 | SS3 |
| $\beta_L$ | 19.82 | 0.01 | 0.00 | 0.00 | 0.01 | 0.36 | 0.37 | 0.37 | 0.37 | 0.74 | 0.22 | 1.10 | 0.01 |
| $\beta_S$ | 4.67 | -0.02 | -0.05 | 0.05 | -0.01 | 0.11 | 0.11 | 0.11 | 0.11 | -0.28 | 0.94 | 0.78 | -0.56 |
| $\sigma_e^2$ | 12.83 | 3.17 | 1.75 | 1.06 | 1.33 | 0.95 | 0.94 | 0.91 | 0.90 | 1.33 | -0.39 | 1.10 | 1.05 |
| $\sigma_L^2$ | 19.85 | -2.34 | 1.06 | 2.65 | 2.46 | 2.81 | 2.86 | 2.88 | 2.85 | 1.34 | -0.35 | 0.02 | -0.70 |
| $\sigma_S^2$ | 1.53 | -2.53 | 0.24 | 2.49 | 2.38 | 0.24 | 0.26 | 0.25 | 0.25 | -1.73 | 1.27 | -0.38 | 0.60 |
| $\rho$ | 0.56 | **10.72** | 2.02 | -0.32 | 0.25 | 0.12 | 0.12 | 0.11 | 0.12 | -0.70 | -1.00 | 0.86 | 0.73 |

*Note: SD is the Bayesian standard deviation.*

Table 3
*Parameter estimates of the Gompertz model in analyzing ECLS-K data*

| | Estimates | | | | SD | | | | Geweke statistic | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Par | IW | SS1 | SS2 | SS3 | IW | SS1 | SS2 | SS3 | IW | SS1 | SS2 | SS3 |
| $\gamma$ | -4.45 | 0.00 | 0.01 | 0.00 | 0.10 | 0.02 | 0.02 | 0.03 | 0.73 | 0.71 | -1.22 | 0.82 |
| $\beta_1$ | 5.24 | 1.53 | 1.52 | 1.53 | 0.10 | 0.03 | 0.03 | 0.03 | -0.73 | -0.65 | 1.21 | -0.71 |
| $\beta_2$ | -47.32 | 0.42 | 0.42 | 0.42 | 1.38 | 0.01 | 0.01 | 0.01 | 1.05 | 0.31 | -1.90 | 0.13 |
| $\beta_3$ | 95.53 | 1.42 | 1.45 | 1.42 | 1.41 | 0.07 | 0.07 | 0.08 | -0.17 | 0.72 | -0.49 | 1.02 |
| $\sigma_e^2$ | 0.21 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 1.74 | -0.66 | 0.53 | -0.63 |
| $\sigma_1^2$ | 0.01 | 0.02 | 0.02 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | -1.70 | -0.99 | -1.55 | -0.35 |
| $\sigma_2^2$ | 0.90 | 0.01 | 0.01 | 0.01 | 0.85 | 0.00 | 0.00 | 0.00 | -0.37 | 0.69 | 0.11 | 0.81 |
| $\sigma_3^2$ | 1.40 | 0.31 | 0.32 | 0.32 | 1.77 | 0.03 | 0.03 | 0.04 | -0.31 | 1.79 | 0.59 | 0.83 |
| $\rho_1$ | 0.00 | 0.74 | 0.70 | 0.73 | 0.12 | 0.09 | 0.08 | 0.09 | 0.62 | -0.37 | -1.40 | 0.22 |
| $\rho_2$ | 0.00 | -0.51 | -0.50 | -0.50 | 0.12 | 0.07 | 0.07 | 0.08 | -1.16 | 0.47 | 1.75 | -0.57 |
| $\rho_3$ | 0.11 | -0.89 | -0.89 | -0.89 | 0.53 | 0.03 | 0.02 | 0.03 | -0.74 | 0.11 | -1.15 | -0.47 |

*Note: SD is the Bayesian standard deviation.*

Table 4

*The parameter estimates of fixed effects when $\sigma_e^2 = 20$ and $N = 50$.*

| | | | BIAS | | | | DCR | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_e^2$ | $\rho$ | Par | IW | SS1 | SS2 | SS3 | IW | SS1 | SS2 | SS3 |
| 1 | 0 | $\beta_L$ | -0.06 | -0.06 | -0.07 | -0.07 | -0.02 | 0.00 | 0.00 | 0.00 |
| | | $\beta_S$ | 0.21 | 0.21 | 0.22 | 0.21 | -0.01 | **-0.03** | -0.01 | -0.01 |
| | 0.5 | $\beta_L$ | -0.18 | -0.18 | -0.18 | -0.18 | -0.02 | 0.00 | 0.00 | 0.00 |
| | | $\beta_S$ | -0.20 | -0.20 | -0.19 | -0.20 | 0.02 | 0.02 | 0.02 | 0.02 |
| | 0.8 | $\beta_L$ | 0.18 | 0.18 | 0.18 | 0.18 | 0.00 | 0.01 | 0.01 | 0.01 |
| | | $\beta_S$ | 0.24 | 0.24 | 0.24 | 0.24 | 0.00 | -0.01 | 0.00 | 0.01 |
| 3 | 0 | $\beta_L$ | 0.24 | 0.24 | 0.23 | 0.23 | -0.02 | -0.01 | 0.00 | 0.00 |
| | | $\beta_S$ | 0.04 | 0.04 | 0.05 | 0.05 | **-0.03** | -0.01 | -0.01 | -0.01 |
| | 0.5 | $\beta_L$ | 0.12 | 0.12 | 0.11 | 0.12 | **-0.03** | -0.01 | -0.01 | 0.00 |
| | | $\beta_S$ | -0.18 | -0.17 | -0.16 | -0.17 | 0.01 | 0.01 | 0.01 | 0.01 |
| | 0.8 | $\beta_L$ | -0.43 | -0.45 | -0.45 | -0.45 | -0.01 | 0.00 | 0.00 | 0.00 |
| | | $\beta_S$ | 0.06 | 0.05 | 0.05 | 0.05 | -0.02 | -0.02 | -0.01 | -0.02 |
| 5 | 0 | $\beta_L$ | 0.36 | 0.36 | 0.35 | 0.35 | -0.02 | 0.00 | 0.01 | 0.00 |
| | | $\beta_S$ | 0.16 | 0.15 | 0.15 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 0.5 | $\beta_L$ | -0.32 | -0.32 | -0.33 | -0.32 | -0.02 | -0.01 | 0.00 | 0.00 |
| | | $\beta_S$ | -0.29 | -0.30 | -0.29 | -0.30 | 0.01 | 0.02 | 0.02 | 0.02 |
| | 0.8 | $\beta_L$ | -0.11 | -0.12 | -0.13 | -0.12 | 0.01 | 0.02 | 0.02 | 0.02 |
| | | $\beta_S$ | 0.07 | 0.05 | 0.06 | 0.04 | 0.01 | 0.02 | 0.02 | 0.02 |

Table 5
*Estimates of $\sigma_e^2$ when its true value is 20*

| | | | BIAS | | | | DCR | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | $\sigma_S^2$ | $\rho$ | IW | SS1 | SS2 | SS3 | IW | SS1 | SS2 | SS3 |
| | | 0 | *8.38* | *8.81* | 4.73 | 4.81 | **-0.08** | **-0.08** | -0.01 | -0.01 |
| | 1 | 0.5 | 4.69 | 3.60 | 1.78 | 1.81 | -0.01 | -0.01 | -0.01 | 0.00 |
| | | 0.8 | 1.74 | 0.77 | -0.24 | -0.24 | 0.00 | -0.01 | -0.01 | 0.00 |
| | | 0 | **14.54** | 7.36 | 5.25 | 5.31 | **-0.08** | -0.02 | 0.00 | 0.00 |
| 50 | 3 | 0.5 | **10.91** | 4.16 | 3.08 | 3.11 | -0.01 | 0.02 | 0.02 | 0.02 |
| | | 0.8 | 3.57 | -0.33 | -0.92 | -0.92 | 0.00 | -0.01 | -0.01 | -0.01 |
| | | 0 | **15.36** | *7.04* | *5.07* | 5.14 | -0.10 | 0.00 | 0.00 | 0.00 |
| | 5 | 0.5 | **12.63** | 4.05 | 3.07 | 3.04 | **-0.05** | -0.01 | 0.00 | 0.00 |
| | | 0.8 | *6.91* | 1.78 | 1.20 | 1.20 | -0.02 | 0.01 | 0.00 | 0.00 |
| | | 0 | *6.11* | *5.82* | 3.81 | 3.85 | -0.02 | **-0.03** | -0.01 | -0.01 |
| | 1 | 0.5 | 3.14 | 1.94 | 1.11 | 1.14 | 0.01 | 0.01 | 0.02 | 0.02 |
| | | 0.8 | 0.14 | -0.58 | -1.04 | -1.02 | -0.01 | -0.02 | -0.02 | -0.02 |
| | | 0 | *6.75* | 3.69 | 2.95 | 2.94 | -0.04 | -0.02 | -0.02 | -0.02 |
| 100 | 3 | 0.5 | *7.15* | 2.58 | 2.02 | 2.01 | **-0.06** | -0.01 | -0.01 | -0.02 |
| | | 0.8 | 2.95 | 0.08 | -0.27 | -0.25 | 0.01 | 0.02 | 0.02 | 0.02 |
| | | 0 | *5.51* | 2.84 | 2.21 | 2.26 | **-0.03** | 0.00 | 0.00 | 0.00 |
| | 5 | 0.5 | *7.76* | 2.41 | 1.90 | 1.93 | -0.10 | 0.01 | 0.01 | 0.01 |
| | | 0.8 | 4.87 | 0.96 | 0.67 | 0.67 | -0.02 | 0.01 | 0.01 | 0.01 |
| | | 0 | 3.12 | 2.51 | 1.81 | 1.79 | -0.02 | -0.01 | -0.01 | -0.01 |
| | 1 | 0.5 | 2.36 | 1.31 | 0.86 | 0.83 | -0.02 | -0.01 | -0.01 | -0.01 |
| | | 0.8 | 0.00 | -0.46 | -0.76 | -0.76 | 0.01 | 0.01 | 0.01 | 0.01 |
| | | 0 | 2.48 | 1.70 | 1.38 | 1.40 | 0.00 | 0.00 | 0.01 | 0.00 |
| 200 | 3 | 0.5 | 4.30 | 1.73 | 1.44 | 1.45 | -0.04 | 0.00 | -0.01 | -0.01 |
| | | 0.8 | 2.21 | 0.07 | -0.14 | -0.15 | 0.00 | 0.01 | 0.01 | 0.02 |
| | | 0 | 1.61 | 1.01 | 0.76 | 0.77 | -0.02 | -0.01 | -0.01 | -0.02 |
| | 5 | 0.5 | 3.92 | 1.48 | 1.23 | 1.23 | **-0.04** | 0.01 | 0.00 | 0.01 |
| | | 0.8 | 3.18 | 0.28 | 0.10 | 0.11 | -0.02 | 0.00 | 0.00 | 0.00 |

*Note. A bold number is either a significant bias(BIAS>10%) or a discrepancy of a bad coverage rate; an italic number represents a moderate bias.*

Table 6

*BIAS on the estimates of* $\mathbf{\Psi}$ *when* $\sigma_e^2 = 20$ *and* $\sigma_S^2 = 1, 3, 5$

| $\sigma_S^2$ | | | 1 | | | | 3 | | | | 5 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N | $\rho$ | | IW | SS1 | SS2 | SS3 | IW | SS1 | SS2 | SS3 | IW | SS1 | SS2 | SS3 |
| 50 | 0 | $\sigma_L^2$ | **-11.25** | 4.15 | *9.83* | *9.60* | **-24.21** | -3.74 | 4.51 | 4.09 | **-22.64** | -0.71 | *7.82* | *7.23* |
| | | $\sigma_S^2$ | *-7.60* | **-18.29** | *5.73* | *5.32* | **-13.56** | -1.91 | 4.08 | 3.82 | *-6.37* | 3.78 | *8.38* | *7.91* |
| | | $\rho$ | **24.64** | **17.97** | **10.90** | **11.00** | **34.90** | **13.23** | *8.76* | *8.98* | **30.34** | *9.50* | *6.25* | *6.43* |
| | 0.5 | $\sigma_L^2$ | -4.49 | **13.17** | **16.87** | **16.78** | **-17.52** | 2.78 | *9.21* | *8.87* | **-18.59** | 3.89 | **10.56** | **10.32** |
| | | $\sigma_S^2$ | *7.50* | 0.75 | **15.09** | **15.09** | -0.47 | *8.62* | **13.81** | **13.74** | -4.32 | 4.06 | *8.06* | *8.13* |
| | | $\rho$ | **20.40** | -0.15 | *-6.94* | *-7.16* | **46.08** | 2.27 | -2.40 | -2.32 | **48.29** | 1.53 | -2.05 | -2.05 |
| | 0.8 | $\sigma_L^2$ | 1.47 | **17.00** | **21.34** | **20.60** | *-5.81* | *8.72* | **14.59** | **14.33** | *-8.39* | *6.60* | **12.81** | **12.49** |
| | | $\sigma_S^2$ | **23.80** | **13.99** | **25.29** | **25.57** | *6.32* | *9.26* | **14.15** | **14.19** | *5.55* | *9.25* | **13.34** | **13.27** |
| | | $\rho$ | *-8.86* | **-19.06** | **-21.94** | **-21.87** | *6.76* | **-11.04** | **-12.55** | **-12.55** | ***10.92*** | *-8.33* | *-9.38* | *-9.41* |
| 100 | 0 | $\sigma_L^2$ | **-11.47** | -4.06 | -0.68 | -0.69 | **-10.04** | -0.54 | 2.92 | 2.84 | **-10.84** | -2.09 | 1.27 | 1.06 |
| | | $\sigma_S^2$ | **-13.64** | **-14.65** | -3.02 | -3.12 | *-6.26* | 0.05 | 2.88 | 2.92 | -1.71 | 2.77 | 4.83 | 4.72 |
| | | $\rho$ | **21.52** | **18.41** | **12.29** | **12.39** | **16.22** | *6.95* | *5.06* | *5.05* | **11.11** | 4.26 | 2.93 | 3.02 |
| | 0.5 | $\sigma_L^2$ | -3.55 | 4.98 | *7.67* | *7.44* | **-11.00** | 1.25 | 4.30 | 4.19 | **-12.95** | 0.24 | 3.40 | 3.19 |
| | | $\sigma_S^2$ | -0.40 | 0.08 | *6.66* | *6.55* | -4.99 | 1.92 | 4.46 | 4.51 | -3.61 | 1.96 | 3.89 | 3.93 |
| | | $\rho$ | **21.59** | *8.19* | 2.16 | 2.31 | **32.96** | 2.85 | 0.14 | 0.11 | **34.93** | 4.89 | 2.96 | 3.08 |
| | 0.8 | $\sigma_L^2$ | 0.19 | *7.44* | *9.82* | *9.53* | -4.40 | 4.20 | *7.08* | *7.02* | *-6.87* | 2.93 | *5.86* | *5.72* |
| | | $\sigma_S^2$ | **13.46** | *9.17* | **14.30** | **14.53** | 1.31 | 3.90 | *6.27* | *6.21* | 0.06 | 2.85 | 4.77 | 4.72 |
| | | $\rho$ | -4.06 | *-9.28* | **-11.32** | **-11.23** | *7.98* | -4.33 | *-5.21* | *-5.14* | **10.82** | -2.83 | -3.44 | -3.41 |
| 200 | 0 | $\sigma_L^2$ | -4.20 | -0.71 | 1.27 | 1.21 | -3.78 | -0.34 | 1.28 | 1.21 | -2.69 | 0.46 | 1.99 | 1.90 |
| | | $\sigma_S^2$ | *-7.76* | -4.84 | -0.32 | -0.21 | -2.74 | -0.26 | 1.09 | 0.97 | -1.92 | -0.14 | 0.81 | 0.84 |
| | | $\rho$ | **11.47** | *8.86* | *6.07* | *6.11* | 4.50 | 2.25 | 1.43 | 1.50 | 3.20 | 1.76 | 1.21 | 1.23 |
| | 0.5 | $\sigma_L^2$ | -4.10 | 0.63 | 2.09 | 2.05 | *-7.10* | -0.50 | 1.00 | 0.92 | *-7.58* | -1.29 | 0.20 | 0.14 |
| | | $\sigma_S^2$ | -3.57 | -0.76 | 2.58 | 2.85 | -4.17 | 0.19 | 1.44 | 1.45 | -1.78 | 1.23 | 2.16 | 2.17 |
| | | $\rho$ | **19.68** | *8.45* | 4.86 | 4.61 | **21.89** | 4.25 | 2.84 | 2.92 | **17.93** | 3.40 | 2.49 | 2.50 |
| | 0.8 | $\sigma_L^2$ | -0.21 | 3.49 | 4.77 | 4.70 | -4.04 | 1.47 | 2.93 | 2.89 | *-5.89* | 0.64 | 2.17 | 2.09 |
| | | $\sigma_S^2$ | *8.35* | *6.86* | *9.63* | *9.77* | -0.49 | 1.93 | 3.12 | 3.20 | -1.40 | 0.90 | 1.86 | 1.92 |
| | | $\rho$ | -3.19 | *-6.42* | *-7.85* | *-7.82* | *6.87* | -2.00 | -2.58 | -2.57 | *8.04* | -1.76 | -2.14 | -2.10 |

*Note: Bold numbers indicate significant biases and italic numbers represent moderate biases.*

Table 7

*Discrepancy of the coverage rate of the estimates of* $\mathbf{\Psi}$ *with* $\sigma_e^2 = 20$

| $\sigma_S^2$ | | | 1 | | | | 3 | | | | 5 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N | $\rho$ | | IW | SS1 | SS2 | SS3 | IW | SS1 | SS2 | SS3 | IW | SS1 | SS2 | SS3 |
| 50 | 0 | $\sigma_L^2$ | **-0.04** | 0.00 | 0.00 | 0.00 | **-0.14** | **-0.03** | -0.01 | -0.01 | **-0.13** | **-0.03** | 0.00 | 0.00 |
| | | $\sigma_S^2$ | **0.03** | **-0.09** | 0.00 | -0.01 | **-0.03** | 0.00 | 0.02 | 0.01 | **-0.04** | -0.01 | 0.00 | -0.01 |
| | | $\rho$ | **-0.04** | 0.02 | 0.02 | **0.03** | **-0.13** | 0.01 | 0.01 | 0.01 | **-0.15** | -0.01 | -0.01 | -0.01 |
| | 0.5 | $\sigma_L^2$ | **-0.05** | -0.02 | -0.02 | **-0.03** | **-0.07** | 0.00 | 0.00 | 0.00 | **-0.07** | 0.01 | 0.01 | 0.01 |
| | | $\sigma_S^2$ | **0.03** | -0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | $\rho$ | 0.01 | 0.05 | 0.04 | 0.04 | **-0.12** | 0.03 | 0.03 | 0.03 | **-0.19** | 0.02 | 0.02 | 0.01 |
| | 0.8 | $\sigma_L^2$ | -0.01 | 0.01 | 0.00 | 0.00 | -0.01 | 0.00 | 0.00 | -0.02 | **-0.03** | -0.01 | -0.01 | -0.01 |
| | | $\sigma_S^2$ | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 | -0.01 | -0.01 | -0.01 | 0.01 | 0.01 | -0.01 | -0.01 |
| | | $\rho$ | **0.05** | **0.04** | **0.03** | **0.03** | **0.04** | **0.03** | **0.03** | **0.03** | -0.01 | **0.04** | **0.04** | **0.03** |
| 100 | 0 | $\sigma_L^2$ | **-0.06** | -0.02 | 0.00 | 0.00 | **-0.06** | -0.01 | -0.01 | -0.01 | **-0.06** | -0.02 | -0.01 | -0.01 |
| | | $\sigma_S^2$ | **-0.03** | **-0.06** | 0.01 | 0.00 | -0.02 | 0.00 | 0.01 | 0.01 | -0.02 | -0.01 | -0.01 | 0.00 |
| | | $\rho$ | **-0.06** | 0.00 | 0.01 | 0.01 | **-0.06** | 0.00 | 0.01 | 0.01 | **-0.05** | -0.02 | -0.02 | -0.02 |
| | 0.5 | $\sigma_L^2$ | **-0.03** | 0.00 | 0.00 | -0.01 | -0.04 | 0.03 | 0.02 | 0.02 | **-0.08** | -0.01 | 0.00 | 0.00 |
| | | $\sigma_S^2$ | 0.02 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | **-0.03** | -0.02 | -0.02 | -0.02 |
| | | $\rho$ | 0.01 | **0.04** | **0.04** | **0.03** | **-0.12** | 0.01 | 0.00 | 0.01 | **-0.18** | 0.00 | 0.00 | 0.00 |
| | 0.8 | $\sigma_L^2$ | 0.01 | 0.00 | -0.01 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | -0.03 | 0.00 | 0.01 | 0.01 |
| | | $\sigma_S^2$ | 0.02 | 0.02 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.01 | 0.00 |
| | | $\rho$ | **-0.05** | **-0.05** | **0.04** | **0.04** | 0.02 | **0.04** | **0.04** | **0.04** | **-0.08** | **0.03** | **0.03** | **0.03** |
| 200 | 0 | $\sigma_L^2$ | **-0.04** | 0.02 | -0.01 | -0.01 | -0.01 | 0.00 | 0.02 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | $\sigma_S^2$ | **-0.03** | -0.02 | -0.01 | -0.01 | -0.02 | -0.01 | -0.01 | -0.01 | -0.02 | -0.01 | 0.00 | -0.01 |
| | | $\rho$ | **-0.06** | **-0.04** | **-0.03** | -0.02 | -0.01 | 0.01 | 0.01 | 0.01 | -0.02 | -0.01 | -0.01 | -0.01 |
| | 0.5 | $\sigma_L^2$ | 0.00 | 0.01 | 0.01 | 0.00 | **-0.05** | -0.01 | -0.02 | **-0.03** | **-0.04** | 0.00 | -0.01 | -0.01 |
| | | $\sigma_S^2$ | **0.03** | 0.02 | 0.02 | 0.02 | -0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | $\rho$ | -0.02 | 0.02 | 0.02 | 0.02 | **-0.09** | 0.01 | 0.01 | 0.01 | **-0.09** | 0.01 | 0.01 | 0.01 |
| | 0.8 | $\sigma_L^2$ | 0.01 | 000 | 0.00 | -0.01 | 0.00 | 0.01 | 0.01 | 0.01 | -0.02 | 0.01 | 0.00 | 0.00 |
| | | $\sigma_S^2$ | 0.00 | 0.00 | 0.00 | -0.01 | 0.01 | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | -0.01 | -0.01 |
| | | $\rho$ | **-0.05** | **0.04** | **0.03** | **0.04** | -0.02 | **0.03** | **0.03** | **0.03** | **-0.09** | 0.00 | 0.00 | 0.00 |

*Note:* *A bold number represents the discrepancy of a coverage rate larger than 0.02 or smaller than -0.02, which corresponds to a coverage rate exceeding the range of [0.93, 0.97].*

Table 8

*Relative biases of parameter estimates in Gompertz model. Bold number represents significant bias.*

| par | true | | SS1 | SS2 | SS3 | IW | SS1 | SS2 | SS3 |
|---|---|---|---|---|---|---|---|---|---|
| | | N=200 | | | | N=500 | | | |
| | | $(\rho_1, \rho_2, \rho_3) = (0, 0, 0)$ | | | | | | | |
| $\gamma$ | 0.15 | **836.23** | *-5.87* | -0.77 | -4.47 | **556.51** | -1.97 | -2.48 | -1.44 |
| $\mu_1$ | 2.80 | **-133.15** | 0.29 | -0.66 | 0.24 | **-93.4** | 0.11 | 0.15 | 0.09 |
| $\mu_2$ | 0.46 | **-511.75** | -0.14 | -0.26 | 0.01 | **-1122.58** | -0.21 | -0.28 | -0.16 |
| $\mu_3$ | 1.56 | **306.00** | -1.06 | -1.32 | -0.75 | **748.05** | -0.31 | -0.42 | -0.19 |
| $\sigma_e^2$ | 0.02 | **832.69** | 1.69 | 1.46 | 1.28 | **2765.59** | 0.62 | 0.44 | 0.5 |
| $\sigma_1^2$ | 0.13 | **86.90** | 0.28 | 3.43 | 1.86 | **-47.91** | -0.18 | 0.64 | 0.41 |
| $\sigma_2^2$ | 0.01 | **4663.90** | *-5.50* | -1.29 | -0.22 | **9910.65** | -2.93 | -1.33 | -0.95 |
| $\sigma_3^2$ | 0.29 | **126.21** | 1.16 | 2.16 | 2.46 | **181.35** | 1.02 | 1.50 | 1.48 |
| $\rho_1$ | 0.00 | **-16.76** | *6.16* | 4.10 | 3.55 | -7.06 | 3.55 | 2.83 | 2.69 |
| $\rho_2$ | 0.00 | **14.63** | -1.07 | -0.62 | -0.58 | -0.11 | -0.58 | 0.15 | -0.39 |
| $\rho_3$ | 0.00 | **-3.31** | 3.92 | 2.50 | 2.20 | 3.47 | 2.44 | 1.91 | 1.86 |
| | | $(\rho_1, \rho_2, \rho_3) = (0.6, -0.5, -0.8)$ | | | | | | | |
| $\gamma$ | 0.15 | **953.22** | -3.45 | 1.27 | -2.20 | **679.59** | -0.93 | 0.48 | -0.19 |
| $\mu_1$ | 2.80 | **-169.41** | 0.31 | -0.61 | 0.25 | **-114.77** | 0.05 | -0.03 | 0.02 |
| $\mu_2$ | 0.46 | **-301.26** | 0.08 | -0.12 | 0.19 | **-928.71** | 0.04 | 0.14 | 0.1 |
| $\mu_3$ | 1.56 | **123.58** | -0.75 | -1.04 | -0.50 | **598.59** | -0.1 | 0.17 | 0.03 |
| $\sigma_e^2$ | 0.02 | **457.11** | 0.71 | 0.62 | 0.52 | **2314.28** | 0.56 | 0.48 | 0.48 |
| $\sigma_1^2$ | 0.13 | **27.93** | 1.05 | 3.01 | 2.03 | **-36.2** | 1.20 | 1.99 | 1.66 |
| $\sigma_2^2$ | 0.01 | **2970.03** | -3.21 | -0.14 | 0.23 | **8493.6** | -2.55 | -1.24 | -1.08 |
| $\sigma_3^2$ | 0.29 | **51.95** | -1.00 | 0.73 | 0.91 | **155.22** | -0.7 | 0.01 | 0.19 |
| $\rho_1$ | 0.60 | **-90.64** | 0.76 | -1.11 | -1.17 | **-108.81** | 1.62 | 0.38 | 0.63 |
| $\rho_2$ | -0.50 | **-105.22** | -2.27 | -3.04 | -2.55 | **-91.17** | -0.24 | -0.87 | -0.39 |
| $\rho_3$ | -0.80 | **-77.38** | -3.17 | -3.00 | -2.80 | **-102.3** | -1.11 | -0.90 | -0.98 |

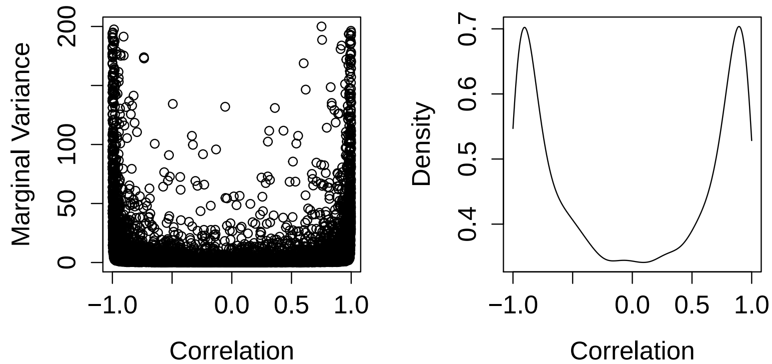
Table 9

*DCR of the parameter estimates in Gompertz model*

| par | true | IW | SS1 | SS2 | SS3 | IW | SS1 | SS2 | SS3 |
|---|---|---|---|---|---|---|---|---|---|
| | | N=200 | | | | N=500 | | | |
| | | $(\rho_1, \rho_2, \rho_3) = (0, 0, 0)$ | | | | | | | |
| $\gamma$ | 0.15 | **-0.59** | -0.01 | 0.01 | 0.01 | **-0.62** | 0.00 | 0.00 | 0.00 |
| $\mu_1$ | 2.80 | **-0.62** | 0.00 | 0.00 | 0.00 | **-0.89** | 0.01 | 0.01 | 0.01 |
| $\mu_2$ | 0.46 | **-0.62** | 0.01 | 0.01 | 0.01 | **-0.86** | 0.00 | 0.00 | 0.00 |
| $\mu_3$ | 1.56 | **-0.54** | -0.02 | -0.02 | -0.01 | **-0.78** | 0.00 | 0.01 | 0.00 |
| $\sigma_e^2$ | 0.02 | **-0.60** | -0.02 | -0.02 | -0.02 | **-0.87** | -0.01 | 0.00 | 0.00 |
| $\sigma_1^2$ | 0.13 | **-0.36** | 0.00 | -0.01 | -0.01 | **-0.73** | 0.00 | 0.00 | 0.00 |
| $\sigma_2^2$ | 0.01 | **-0.95** | -0.02 | -0.01 | -0.01 | **-0.83** | -0.01 | -0.02 | -0.01 |
| $\sigma_3^2$ | 0.29 | 0.02 | -0.01 | 0.00 | -0.01 | **0.05** | 0.00 | -0.01 | -0.02 |
| $\rho_1$ | 0.00 | **-0.32** | 0.00 | 0.01 | 0.00 | **-0.05** | -0.02 | -0.01 | -0.01 |
| $\rho_2$ | 0.00 | **-0.05** | 0.00 | 0.00 | 0.00 | **0.05** | 0.01 | 0.00 | 0.00 |
| $\rho_3$ | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.02 | -0.02 | -0.02 | -0.02 |
| | | $(\rho_1, \rho_2, \rho_3) = (0.6, -0.5, -0.8)$ | | | | | | | |
| $b_0$ | 0.15 | **-0.67** | 0.00 | 0.00 | -0.01 | **-0.88** | -0.01 | 0.00 | -0.01 |
| $\mu_1$ | 2.80 | **-0.69** | -0.01 | -0.01 | -0.01 | **-0.93** | -0.01 | 0.00 | 0.00 |
| $\mu_2$ | 0.46 | **-0.73** | -0.01 | -0.01 | -0.01 | **-0.92** | 0.00 | 0.00 | 0.00 |
| $\mu_3$ | 1.56 | **-0.62** | 0.01 | 0.01 | 0.00 | **-0.88** | 0.00 | -0.01 | 0.00 |
| $\sigma_e^2$ | 0.02 | **-0.57** | 0.00 | 0.00 | 0.01 | **-0.92** | -0.01 | -0.02 | 0.00 |
| $\sigma_1^2$ | 0.13 | **-0.11** | 0.00 | 0.00 | 0.00 | **-0.85** | -0.01 | -0.01 | -0.02 |
| $\sigma_2^2$ | 0.01 | **-0.95** | -0.01 | 0.00 | -0.01 | **-0.90** | 0.00 | -0.03 | -0.01 |
| $\sigma_3^2$ | 0.29 | **-0.18** | -0.01 | -0.02 | -0.02 | **-0.73** | -0.01 | 0.00 | -0.01 |
| $\rho_1$ | 0.60 | **-0.93** | 0.02 | 0.01 | 0.01 | **-0.95** | 0.01 | 0.00 | 0.01 |
| $\rho_2$ | -0.50 | **-0.46** | 0.01 | 0.00 | 0.01 | **-0.87** | 0.01 | 0.01 | 0.01 |
| $\rho_3$ | -0.80 | **-0.82** | 0.00 | 0.01 | 0.01 | **-0.85** | -0.01 | -0.01 | -0.01 |

*Note:* DCR *means discrepancy of coverage rate; bolder number means large DCR.*

*Figure 1*. Visualization of the inverse-Wishart distribution IW(2,$\mathbf{I}_{2\times2}$) based on $10,000$ draws




*Note: The left panel is the scatter plot of the marginal variances and correlation coefficients; the right panel is the density plot of correlation coefficients.*